Karlsruher Institut für Technologie

ITAS Institute for Technology Assessment and Systems Analysis

Academy for Responsible Research, Teaching, and Innovation

ARRTI

Designing AI and Digital Technologies for Humanity

The Role of Ethical Literacy

Rafaela Hillerbrand

23.XI.2022 Vienna

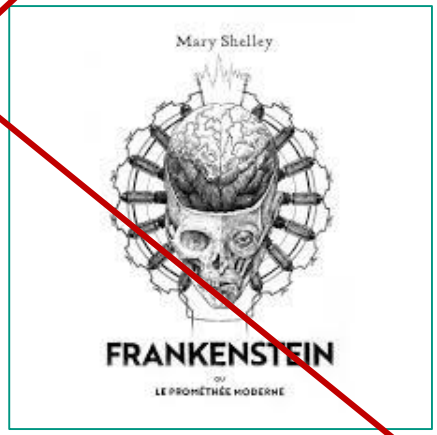
Outline Talks

- **Challenges**
  of (weak) AI (as discussed in the literature/media)

- **Chances**
  of an AI for fostering human values

- **Paths**
  towards ethical literacy

What is AI not (in this talk)?



- Major parts of public and reflective debate focus on some future form of AI

- Urgent for Universities of Technology **now:**
- AI = current or foreseeable AI after the 3rd spring:
  - Machine Learning, (un)supervised,
  - Neuronal nets,
  - ...

What is AI

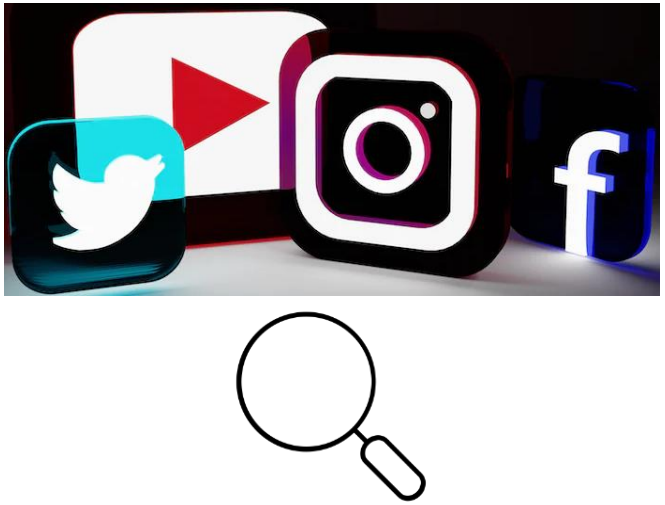
- = a system‘s ability to
  - ➢ correctly interpret external data
  - ➢ learn from such data
  - ➢ use what it learnt to achieve specific goals(s) through flexible adaption

- **Autonomous vehicle**
  - ➢ Knows the difference between a person and a shadow from data provided by a camera
  - ➢ Better distinguishes the difference between a person and a shadow
  - ➢ Steers the vehicle

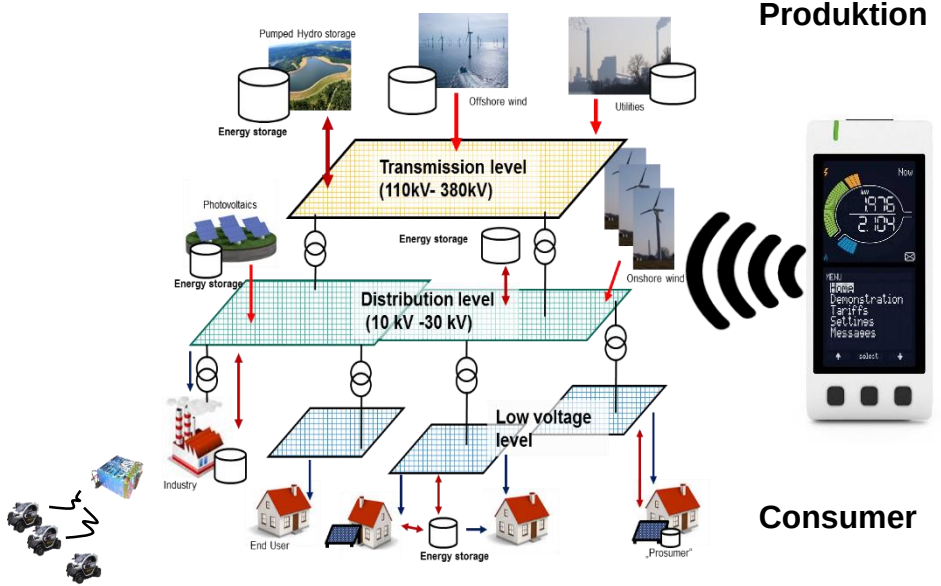
e.g. Kaplan/Haenlein 2019 Business Horizons 62(1), 15

AI is everywhere

Search Algorithms in Social Media Applications



Smart Electrical Grids



From Electricity Grid to Data Grid

Chances

- **Enabling technology** for the energy transition towards more renewable sources
- Smart grids to counterbalance that the intermittent electicity production and usage are decoupled
- AI enables (semi-) autonomous driving
    - Efficiency/energy savings

Risks

- Transport not only of electricity but also of data
- Privacy
- Safety
- (Justice)
- ….
- Introduces unprecedented dynamics in the electricity system
    - typical time spans in the pat covered several decades for infrastructure and power plants

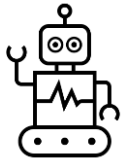
Some of the Risks of AI

- Security
  - From violations of privacy and data integrity
  - ... to deep fakes
- Safety
  - From autonomous weapon systems
  - ... to attacks critical infrastructures like the electrical grid
- Stockmarket volatility
- *Automation* and corresponding job losses
- Biases and hardening of social injustices
- ...

A Classification of the (Ethical) Risks

- Misuse
  - dual use
  - (unwanted) unintended use
  - ....
- Non-intended consequences
  - “accidents“
  - structural risks

This (ethical) **dual nature** of AI is similar to any technology



→ Is there something NEW about AI?

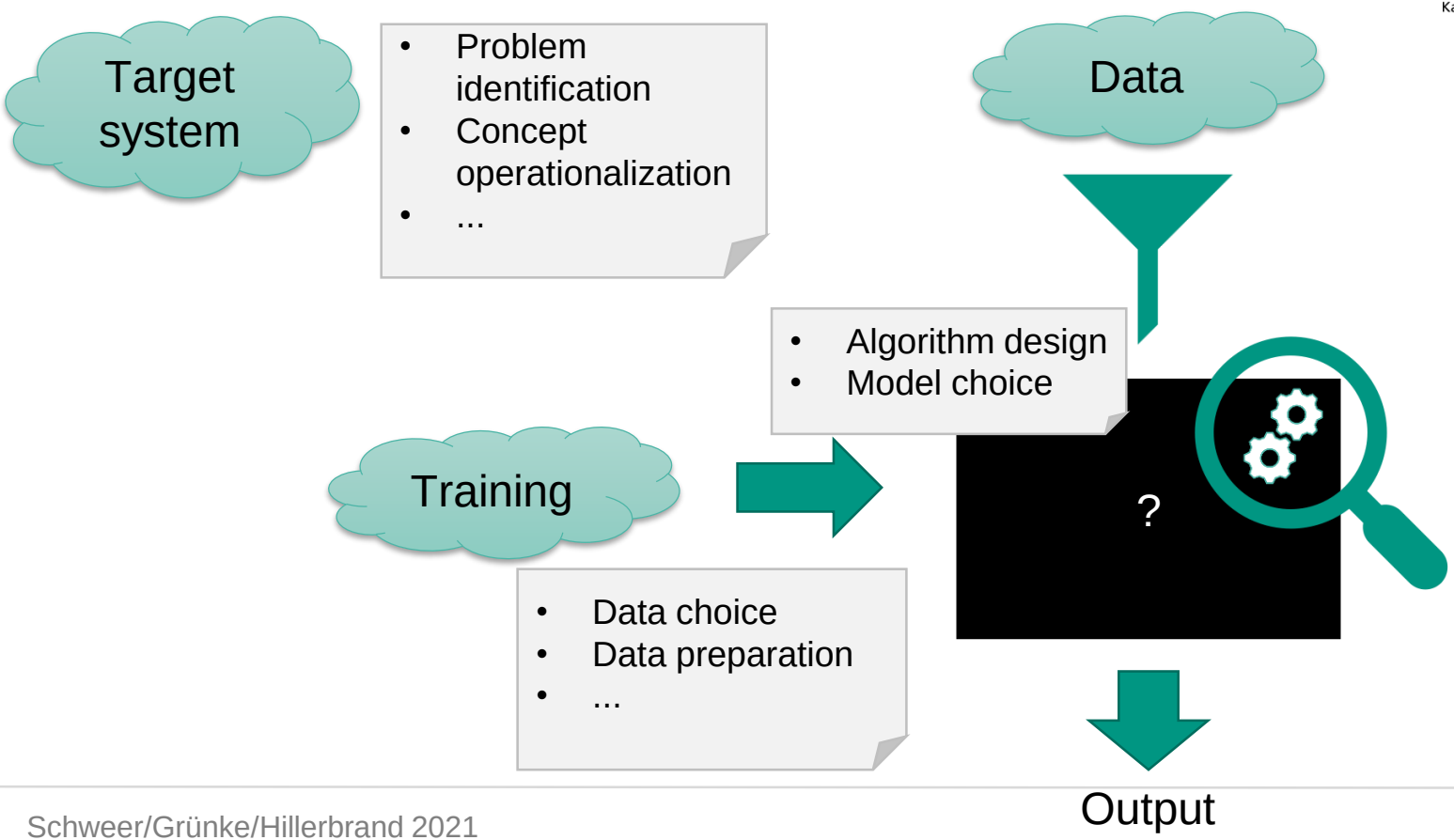
The Dual Nature of AI in Dealing with Complexity

- **AI to reduce complexitites**
- AI used to better understand and deal with risks
- Examples: autonomous driving, natural hazard forecasting, clearance of crime scenes with radioactive exposore, ...

- **AI is a source of complexity**
- AI als „black box“ or as epistemically opaque

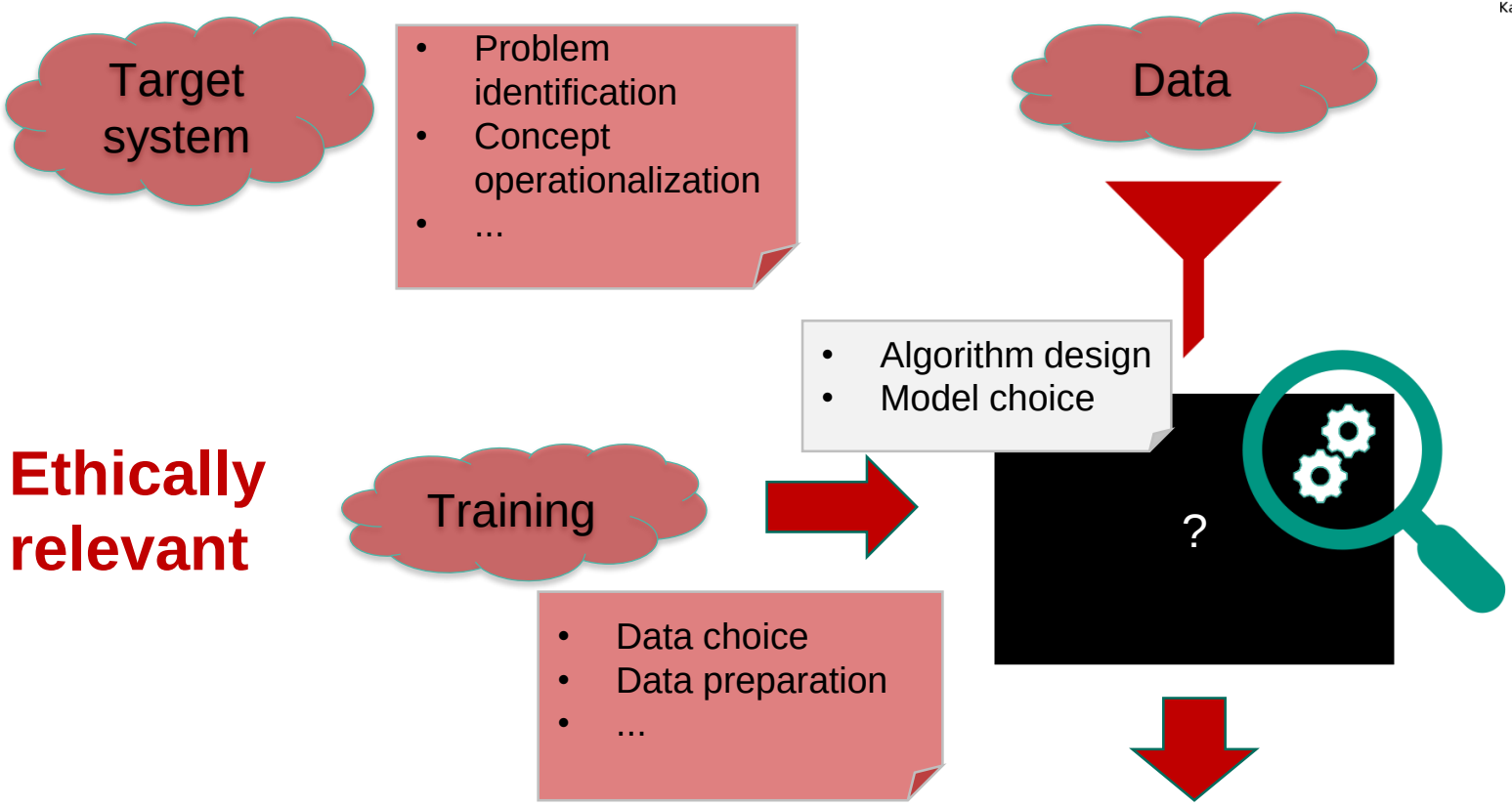
For the user: XAI since 2004

Transparency of AI broader than “Black Box”



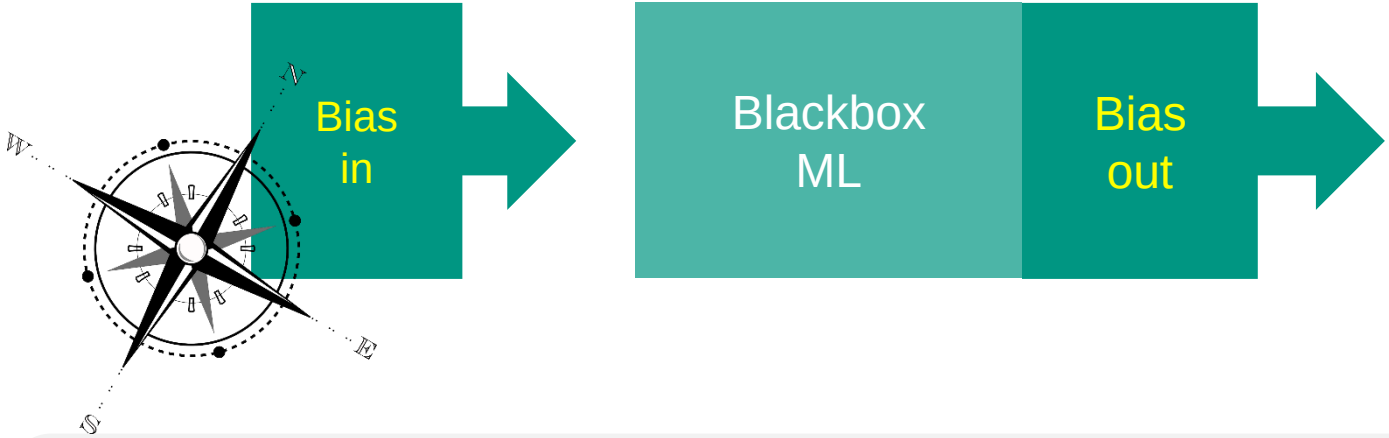
Schweer/Grünke/Hillerbrand 2021

Transparency of AI broader than “Black Box”



Example: AI and Biases/Social Injustices

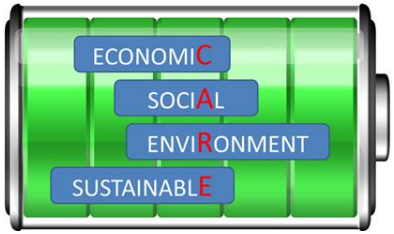
Example: AI and Biases/Social Injustices



- COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions)* as an example of biased decision making due to biased training and learning data set

Potentials of AI

* AI can be used to reduce social inequalities
  * include marginalised groups (e.g. Remesh/UN)
  * reduce the costs of teaching
  * ...
* Instead of an only focus in *threat prevention* use AI to increase out ability to care and make its users more human
  * Future use in Social Life Cycle Assessment
  * …



https://futuringpeace.org/

Outline Ethics & AI

- **Challenges**
  of (weak) AI (as discussed in the literature/media)

- **Chances**
  of an AI for fostering human values

- **Paths**
  for training of engineers to device an AI for humanity

The Janus-Face of Engineering

mathematics

physics

sociology

economics

Application of descriptive scientific and mathematical principles ...

How to shape our world in a normative way ? Which problems are to be solved, ...

law

ethics

...

chemistry

science and technology studies

… shape or re-design the world.

Ethical Literacy

- Competence to
  - identify ethical problems as such
  - develop normative arguments
  - Reasoning about the normative goals of technological systems/artefacts

- In ethical theory this competence is know as a specific virtue, the phronesis

Why Ethics in Engineering Education

- **Complexity** of engineering design and research …
  - Necessitates to train the engineers in ethics (as a casuistic or a satisfactory regulatory framework (ISO norms etc) are out of reach – devil in the (design) detail
  - Makes it difficult to bridge the gap between ethical theories and engineering practice
  - Seems to hinder accepted mid-level principles (like in medical ethics)
  - Makes it hard to assign responsibilities (e.g. due to institutional complexity, cf. Volkswagen emission scandal)
- Technological Artefacts and processes as results of **creative enterprise** are impossible to regulate

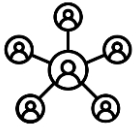
The KIT Academy for Responsible Research Teaching and Innovation – ARRTI

The ARRTI Goal:

Strengthen responsibility in research and innovation (global & intergenerational) & responsibility to innovate

The ARRTI Approach:

Training of **ethical literacy:** ethical competences instead a primary focus on ethical theories and principles



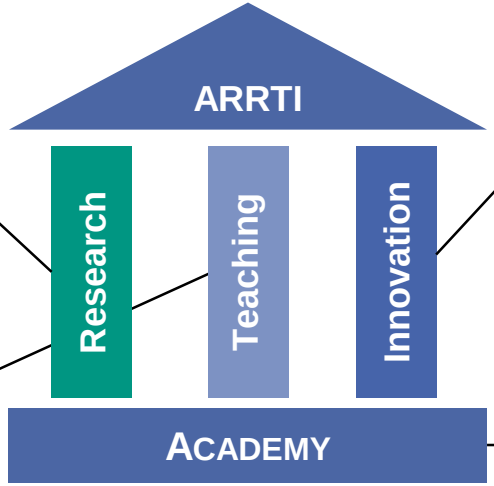
The ARRTI Method:

Ethics as enabling („Ermöglichung“):

- Inter- & transdisciplinary
- Bottom-up
- Free from ideologies

Ways to Ethical Literacy

Courses for PhDs and PostDocs on Good Scientific Practice or Research Ethics; specifc courses for research staff
Challenges



Challenges for Resarch Staff and Students

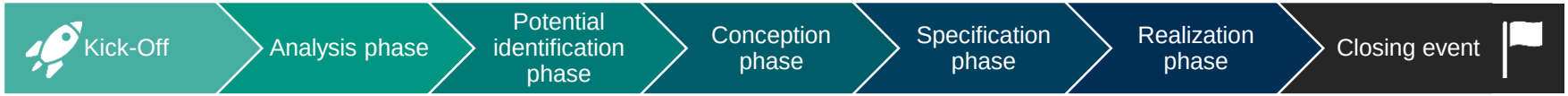
KIT *Sustainable Innovation Challenge*

Seminars, lecture series, blended learning (ARsRef), ...

“Train the Trainers”

Example: Ethics in Integrated Product Development

5 months



1. Lectures

Ethics, Technology and Responsibility

Engineering action and responsibility

2. Workshop

Blending theory and practice

3. Reflection meetings

5 per team (90 min)

4. Lecture series

“Role models in thought and action - Responsible innovation in practice”

5. Reflection meeting

with supervisors

ARRTI in IP – Integrated Product Development

Student voices

„[…] very important to drive forward and advance society through technology and innovation. And that‘s why these **ethical aspects are also very important to me** at this point.“

„[…] the **perception of responsibility** that one has **as an engineer has become a different one**, because you don‘t just build things, you build things for someone, and they should be able to use them safely.“

Summary

- In order to adequately address potential threats of technological development
  - a sensitivity (skill/virtue) is needed to anticipate potential ethical problems
- This (part of an) **ethical literacy** is to be trained in various contexts
  - and enables not only to react to potential threats but to design technological artefacts or processes for ethically good proposes
  - From a *threat perspective* to a an overarching *care perspective* of technology
- AI is not special in this respect,
  - but due to its dual epistemic nature,
  - it is mainly to the developers to address the ethical issues
- → greater necessity to include ethical literacy in engineering curricula

Karlsruhe Institute of Technology

Thank You for Your Attention!

Academy for Responsible Research, Teaching, and Innovation

ARRTI